

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/97888>

Please be advised that this information was generated on 2017-12-06 and may be subject to change.

Evaluation of a Computer-Aided Diagnosis System for Diabetic Retinopathy Screening on Public Data

Clara I. Sánchez,¹ Meindert Niemeijer,^{2,3} Alina V. Dumitrescu,⁴
Maria S. A. Suttorp-Schulten,⁵ Michael D. Abràmoff,^{3,4} and Bram van Ginneken¹

PURPOSE. To evaluate the performance of a comprehensive computer-aided diagnosis (CAD) system for diabetic retinopathy (DR) screening, using a publicly available database of retinal images, and to compare its performance with that of human experts.

METHODS. A previously developed, comprehensive DR CAD system was applied to 1200 digital color fundus photographs (nonmydriatic camera, single field) of 1200 eyes in the publicly available Messidor dataset (Methods to Evaluate Segmentation and Indexing Techniques in the Field of Retinal Ophthalmology (<http://messidor.crihan.fr>). The ability of the system to distinguish normal images from those with DR was determined by using receiver operator characteristic (ROC) analysis. Two experts also determined the presence of DR in each of the images.

RESULTS. The system achieved an area under the ROC curve of 0.876 for successfully distinguishing normal images from those with DR with a sensitivity of 92.2% at a specificity of 50%. These compare favorably with the two experts, who achieved sensitivities of 94.5% and 91.2% at a specificity of 50%.

CONCLUSIONS. This study shows, for the first time, the performance of a comprehensive DR screening system on an independent, publicly available dataset. The performance of the system on this dataset is comparable with that of human experts. (*Invest Ophthalmol Vis Sci.* 2011;52:4866–4871) DOI:10.1167/iovs.10-6633

Diabetic retinopathy (DR) is the most common cause of blindness in the working population of the United States and Europe. DR will become a more important problem worldwide. The World Health Organization (WHO) predicts that the number of patients with diabetes will increase to 366 million in 2030.¹ In patients with diabetes, early diagnosis and treatment have been shown to prevent visual loss and blindness.^{2–4} However, more than 50% of the diabetes pop-

ulation worldwide does not undergo any form of eye examination.⁵ The use of digital photography of the retina examined by expert readers during screening programs has been shown to be both sensitive and specific in the detection of the early signs of diabetic retinopathy.^{6,7} Access to screening services is an increasingly important and pressing issue, especially given the increasing prevalence of diabetes. To increase access to screening, several groups have proposed the use of automated computer systems for determining what screened patients should be seen by an ophthalmologist and what patients can safely return for screening 1 year later.^{5,8,9} These types of automated systems have the potential to reduce the workload for screening ophthalmologists while maintaining a high sensitivity (i.e., above 90%) for the detection of patients with DR.

For automated systems to be applied in clinical practice, they should be evaluated extensively and thoroughly. One of the goals of this evaluation is to show that automated systems can detect DR with a sensitivity comparable to that of a human expert while maintaining a high enough specificity to attain the needed reduction in the ophthalmologist's workload. In addition, evaluation of systems should be performed on independent and, preferably, publicly available data so that different groups can compare the performance of their automated systems on the same set of data. Of additional importance, the performance record of several expert observers on this same dataset should also be available to facilitate the comparison between automated systems and humans.

Currently, two large studies (i.e., involving more than 10,000 examinations) by two research groups have been published.^{5,8} These studies used datasets that the authors expected to be typical of the populations on which their proposed systems would be used. Although internally valid, this approach does not allow external validity (on other populations and datasets) to be determined. Recent work on discriminating between normal and pathologic retinal images¹⁰ has been evaluated with a small subset from a public database. Many more groups have evaluated components of DR screening systems on smaller datasets.¹¹ Recently, more public data for the evaluation of algorithms have become available.^{12–14} The largest publicly available dataset is Messidor, consisting of 1200 macula-centered digital fundus photographs (<http://messidor.crihan.fr>).¹⁴ Although this dataset was obtained in a clinical setting and thus with a distribution of diabetic retinopathy disease severity different from a screening population, it has the advantage of increased external validity, because this distribution is wider, in addition to its public availability.

The purpose of the present study was to apply our comprehensive automated DR screening system to the Messidor dataset and compare its performance with that of two human experts.

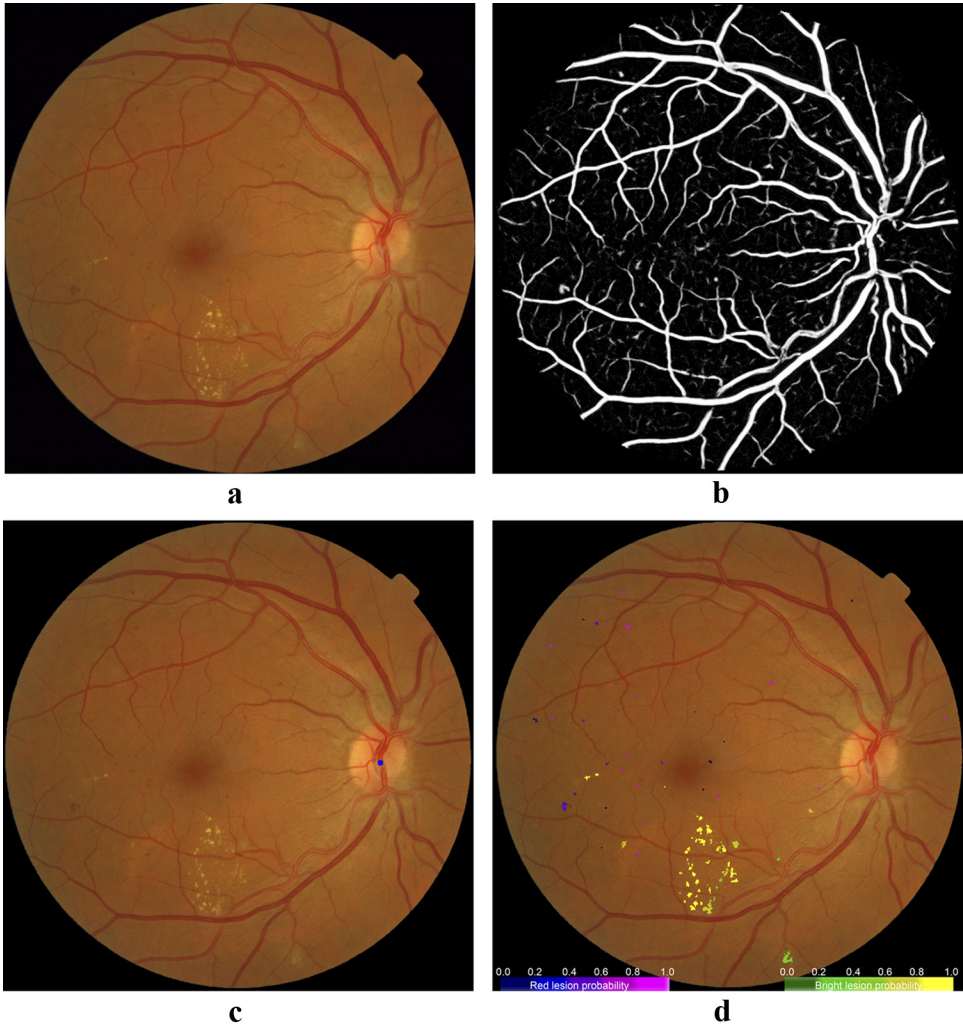
From the ¹Diagnostic Image Analysis Group, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands; the ²University Medical Center Utrecht, Image Sciences Institute Utrecht, Utrecht, The Netherlands; the ³Department of Electrical and Computer Engineering, The University of Iowa, Iowa City, Iowa; the ⁴Department of Ophthalmology and Visual Sciences, The University Of Iowa Hospitals and Clinics, Iowa City, Iowa; and the ⁵Ophthalmology Service, OLVG (Onze Lieve Vrouwe Gasthuis) Hospital, Amsterdam, The Netherlands.

Submitted for publication September 24, 2010; revised February 9, 2011; accepted February 11, 2011.

Disclosure: C.I. Sánchez, None; M. Niemeijer, None; A.V. Dumitrescu, None; M.S.A. Suttorp-Schulten, None; M.D. Abràmoff, None; B. van Ginneken, None

Corresponding author: Clara I. Sánchez, Diagnostic Image Analysis Group, Radboud University Nijmegen Medical Center, Department of Radiology, Postbus 9101, 6500 HB Nijmegen, The Netherlands; c.sanchezgutierrez@rad.umcn.nl.

FIGURE 1. Examples of the outputs of the proposed CAD system. **(a)** Original image from the Messidor database (filename: 20051020_57566_0100_PP.tif), kindly provided by the Messidor program partners (<http://messidor.crihan.fr/download-en.php>). The quality-verification module automatically assigned a probability of 0.98 that the image would have good quality. **(b)** Output of the automatic vessel segmentation module. The image shows the obtained pixel probability map indicating the likelihood of the pixel to belong to a vessel. *White*: higher probability. **(c)** Output of the automatic optic disc detection module. *Blue spot*: the obtained location within the image with the highest probability of being the optic disc center. **(d)** Outputs of the automatic red and bright lesion-detection modules. Each candidate is assigned a value indicating the probability of being a true lesion. The color scales represent the range of values for the red and bright lesion probability.



METHODS

Data

The Messidor database¹⁴ was established to facilitate studies on computer-aided diagnosis of DR. The database consists of 1200 color fundus images of the posterior pole. The included patients were randomly chosen among the diabetic patients from the ophthalmology departments involved in the Messidor project.¹⁴ An example of an image from the Messidor database is shown in Figure 1a. The images were acquired in three different ophthalmology departments, 400 images in each department, using a nonmydriatic digital retinal camera (TRC NW5; TopCon, Tokyo, Japan) with 45° field of view. Eight hundred images were acquired with pupil dilation (one drop of tropicamide at 0.5%) and 400 images without dilation. Image sizes were 1440 × 960 in 588 images, 2240 × 1488 in 400 images, and 2304 × 1536 in 212 images. All the images were saved in uncompressed TIFF format.

For each image, two diagnoses, retinopathy grade, and risk of macular edema, have been provided with the dataset. These diagnoses were obtained by medical experts according to the grading schemes shown in Table 1 (Erginay A, et al. *IOVS* 2008;49:ARVO E-Abstract 2137). The diagnoses were considered to be the reference standard for the performance analysis in our work. According to the reference standard, a total of 546 images were classified as normal and 654 as presenting signs of DR, specifically 153 with retinopathy grade 1, 247 with retinopathy grade 2, and 254 with retinopathy grade 3. In addition, 974 images do not show risk of macular edema; whereas 75 and

151 images presented risk grades 1 and 2 for macular edema, respectively. Information about patients was removed to ensure patient privacy.

Computer-Aided Diagnosis System for Diabetic Retinopathy Screening

The proposed computer-aided diagnosis (CAD) system analyzes a patient's examination to identify lesions associated with DR and assigns

TABLE 1. Grading Schemes Proposed for Retinopathy Grade and Risk of Macular Edema¹⁴

Grade	Description
Retinopathy grade	
0*	($\mu A = 0$) and ($H = 0$)
1	($0 < \mu A \leq 5$) and ($H = 0$)
2	[$(5 < \mu A < 15)$ and ($0 < H < 5$)] and ($NV = 0$)
3	($\mu A \geq 15$) or ($H \geq 5$) OR ($NV > 0$)
Risk of macular edema	
0*	No visible hard exudates
1	Shortest distance between macula and hard exudates > one papilla diameter
2	Shortest distance between macula and hard exudates \leq one papilla diameter

μA , microaneurysms; H, hemorrhages; NV, neovascularization.
* Normal.

each examination a probability between 0 and 1, with 1 indicating that the examination should be referred to an ophthalmologist. A patient is deemed referable if the examination contains DR lesions or the examination is ungradable due to low quality. To accomplish this, the proposed system consists of various modules responsible for the following tasks (shown in Fig. 1).

Preprocessing. Before finding anatomic landmarks and lesions within the image, its field of view (FOV) is detected by finding the optimal FOV template among a predefined group of templates that match the image. The image is then resized to have an FOV with a standardized diameter of 650 pixels independent of the image resolution.¹⁵

Quality Verification. This module determines the quality level of the image. The technique relies on the assumption that an image of sufficient quality should contain particular image structures—namely, the vasculature, the optic disc (OD), and the background, according to

a certain predefined distribution. A compact representation of the image structures is obtained applying a Gaussian filter bank (GFB) to the image and clustering the outputs. One cluster represents one structure. The distribution of the image structures within the image is then represented by means of a histogram with one bin per cluster. Using this histogram together with histograms of the R, G, and B color planes as features, a support vector machine is trained to assess the image quality. The output of this module is a probability per image indicating the likelihood the quality of the image is normal. It should be noted that the output of this module is not used to discard images of low quality, only to analyze the quality level.¹⁶

Vessel Segmentation. The vasculature is one of the most important anatomic structures in retinal images. Vessel segmentation is necessary to distinguish small vessels from red lesions and as an aid for the identification of other anatomic landmarks, such as the OD. A pixel probability map indicating the likelihood that the pixel belongs to a

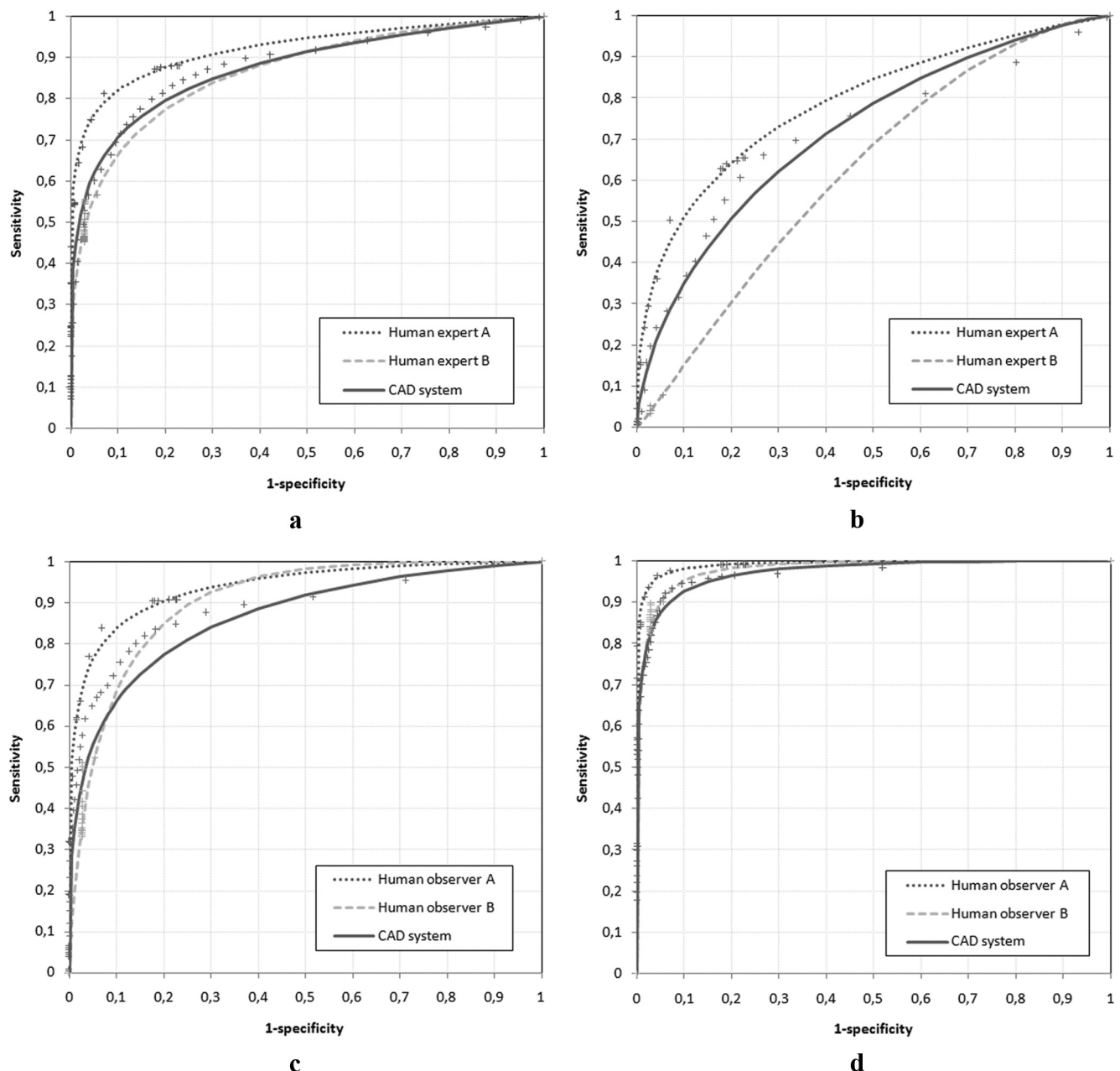


FIGURE 2. Fitted ROC curves for the human experts and the CAD system: (a) Normal and abnormal ROC curves; (b) Normal/grade 1 ROC curve; (c) normal/grade 2 ROC curve; and (d) normal/grade 3 ROC curve.

vessel is obtained as output by means of pixel classification using GFB features and a supervised classifier.¹⁷

OD Detection. The OD is another important anatomic structure. The identification of this element is necessary to prevent erroneously detected bright lesions within the OD. The OD is identified calculating a regression rule between its center location and a group of features based on intensity, vessel orientation, and density. The output of the module is a location within the image with the highest probability that it is the OD center.¹⁸

Red Lesion Detection. Red lesions, comprising microaneurysms, hemorrhages, and vascular abnormalities, are important signs of DR and their detection is therefore of paramount importance for a DR screening system. Potential red lesion locations are identified by using a hybrid approach based on mathematical morphology, specifically designed for smaller candidates, and a supervised pixel classification using GFB features, for the detection of larger red lesions. The detected candidates are then assigned a probability of being a true red lesion, using a supervised classifier and a group of features describing the candidate shape, structure, color, and contrast.¹⁹

Bright Lesion Detection. Bright lesions, such as exudates, cotton wool spots, or drusen, are frequently encountered in a DR population screening. Only the first two are associated with DR. Similar to red lesion detection, a supervised pixel classification is first performed to obtain candidates that may be bright lesions. A probability that each candidate is a true bright lesion is then obtained by means of supervised classification using a group of candidate features, such as shape, contrast, color, and distance to the nearest red lesion.²⁰

The outputs of the different modules must be combined to obtain a final decision about the patient's examination. To accomplish that, a group of features based on the diverse outputs of the aforementioned modules are calculated, such as the quality likelihood or the highest likelihood of red or bright lesions in the examination. These features are given as input to a k nearest-neighbor (k NN) classifier. This classifier was trained on an independent training set not used for any other purpose in this research. The output of this classifier is a per-examination probability indicating the likelihood that the examination would be referred to an ophthalmologist.¹⁵

Observer Study

To compare the performance of the CAD system with that of experts, a general ophthalmologist and a retinal specialist, with 4 and 20 years of DR screening experience, respectively, manually analyzed the images in the Messidor database. Both experts have experience with digital and real-time examinations. The specialists were asked to provide a value between 0 and 100 for each image, indicating the probability of the presence of DR in the image. In addition, the specialists evaluated the retinopathy grade and the risk of macular edema, according to the grading scheme shown in Table 1. The images were revised in different sessions depending on the readers' availability. The resampling images were displayed in a LCD screen without any calibration and with the ability to zoom and pan.

Data Analysis

The performances of the CAD and the two experts were compared separately with the reference standard. For this purpose, ROCKIT software²¹ was used to analyze the outcomes of the CAD and the experts. Using the raw output data as input, the software applied a maximum-likelihood estimation to fit a binormal receiver operating characteristic (ROC) curve.²² The area under the ROC curve, A_z , was used as a measure of the system or human performance, and a univariate z -score test was performed to compare the performance of the CAD system and the experts. Overall agreement between the specialists and the reference standard was calculated using weighted κ statistics (SPSS, ver. 17.0.0; SPSS, Chicago, IL).

We performed four experiments for both the CAD system and the experts, obtaining four ROC curves:

- Normal/abnormal ROC curve: All the images in the MESSIDOR database were used.
- Normal/grade 1 ROC curve: Only normal images and images with retinopathy grade 1 according to the reference standard were used.
- Normal/grade 2 ROC curve: Only normal images and images with retinopathy grade 2 according to the reference standard were used.
- Normal/grade 3 ROC curve: Only normal images and images with retinopathy grade 3 according to the reference standard were used.

These experiments allowed an analysis of the performance for the different retinopathy grades separately.

RESULTS

Human Experts' Performance

The fitted ROC curves for both experts are shown in Figure 2. The areas under the different ROC curves are summarized in Tables 2 to 6 show the contingency tables and the weighted κ agreement between the expert and the reference standard for the assessment of retinopathy grade and risk of macular edema.

CAD System Performance

The fitted ROC curves for the CAD system are shown in Figure 2. The corresponding areas under the different curves are summarized in Table 2. Setting the operating point at 50% specificity on the normal or abnormal ROC curve, a sensitivity of 92.2% was obtained, with a total of 51 abnormal images wrongly classified as normal. The CAD system misclassified 29 images with retinopathy grade 1, 19 with grade 2, and 3 with grade 3, according to the reference standard. Among the 19 misclassified images with grade 2, the experts agreed on only two images as abnormal and both graded seven images as normal. The three misclassified images with grade 3 presented large hemorrhages. Specialist A assessed them as grade 1 or 2 but not grade 3, and specialist B classified two of them as normal. None of the 51 misclassified images presented a risk of

TABLE 2. The A_z under the ROC Curve and the 95% Confidence Interval (CI), as a Measure of the Performance of the Two Experts and the CAD System

	Expert A		Expert B		CAD System	
	A_z	95% CI	A_z	95% CI	A_z	95% CI
Normal/abnormal	0.922	0.902–0.936	0.865	0.789–0.925	0.876*	0.856–0.895
Normal/grade 1	0.789	0.728–0.841	0.623	0.258–0.899	0.721*†	0.673–0.765
Normal/grade 2	0.940	0.971–0.958	0.904	0.839–0.948	0.867*	0.836–0.893
Normal/grade 3	0.992	0.986–0.996	0.981	0.969–0.989	0.973*	0.961–0.982

A_z , area under the ROC curve.

*† CAD system performance is nonsignificantly different ($P > 0.05$) compared with the experts B and A, respectively.

TABLE 3. Confusion Matrices, Weighted κ Agreement, and 95% CI for Retinopathy Grade between the Reference Standard and the Experts

	Expert A				Expert B			
	Grade 0	Grade 1	Grade 2	Grade 3	Grade 0	Grade 1	Grade 2	Grade 3
Grade 0	502	39	5	0	544	0	0	2
Grade 1	72	69	11	1	146	3	3	1
Grade 2	28	64	147	8	117	4	105	21
Grade 3	2	10	84	158	20	15	54	165
	$\kappa = 0.755$				$\kappa = 0.637$			
	95% CI = 0.733–0.780				95% CI = 0.604–0.670			

The columns indicate the observer performance; the rows indicate the reference standard.

macular edema with grade 2. At a specificity of 50%, the experts A and B obtained a sensitivity of 94.5% and 91.2%, respectively, and 77.5% and 87.8% of the images showed a probability higher than 0.95 and 0.8 of having a normal quality, respectively.

DISCUSSION

In this article, a CAD system was evaluated with independent and publicly available data. This kind of evaluation is of paramount importance to obtain reproducible results and allow objective comparison to other DR CAD systems.

The CAD system achieved a performance of the area A_z under the ROC curve of 0.876, similar to the performance obtained by the experts. At a specificity of 50%, the CAD system obtained a sensitivity of 92.2%, comparable to the performance of the experts. With this setting, the system offers a valuable opportunity to reduce the manual burden of grading with a workload reduction of 50%. Among the 51 images misclassified by the system as normal, no images with high risk of macular edema were missed. Only three images with high retinopathy grade were wrongly classified, according to the reference standard. However, the experts did not agree on the grade of those images, one even classifying two of them as having grade 0. CAD offers retinopathy screening programs a fast solution to screen diabetes population. The average time to process one examination is less than 3 minutes for nonmultithreaded software written in C++ running on a 2.66-GHz quad processor (Core 2; Intel, Mountain View, CA).

The performance of the CAD system on the differential retinopathy grading was also comparable to that of the human observers. For the CAD system and for the experts, the differentiation between normal images and images with retinopathy grade 1 according to the reference standard was the most difficult task. In fact, in some screening systems, patients with up to only five microaneurysms are not even referred. When a

distinction between referable (retinopathy grades 2 and 3) and nonreferable (retinopathy grade 0 and 1) images was performed on the Messidor database, areas A_z under the ROC curve of 0.91, 0.94, and 0.92 were obtained by the CAD system and the experts A and B, respectively, with sensitivity of 94.4%, 98.2%, and 97.6% at a specificity of 50%.

The observer study showed that there was low agreement on the grading of retinal images, even when adhering to a strict protocol such as the one proposed in Table 1. These results suggest that human grading is subjective, depending on the reader and his or her experience. An automatic method that could assess automatically the retinopathy grade might be of value for reducing the high interobserver variability of grading. It should be emphasized that the performance of a CAD system cannot exceed the human performance due to the lack of an objective gold standard. The system was trained using the annotations made by specific observers, and its performance therefore depends on the observer's opinion.

The CAD has performance level on the Messidor database similar to those reported in our previous studies using different databases.^{5,15} In these studies, the system achieved an area A_z under the ROC curve of 0.84 and 0.88, with databases of 10,000 and 15,000 examinations, respectively. These comparable performance levels highlight the reliability of the proposed system in the face of data changes, such as screening protocol and number of images per examination and quality. However, it should be noted that the performance may deteriorate in images taken from the screening setting, as they present lower quality than images acquired from clinical settings. In addition, it is reasonable to assume that a higher performance can be obtained in images with a higher resolution. Measuring the CAD performance with respect to the image resolution, areas under the ROC curve of 0.927, 0.914, and 0.935 were obtained for resolutions of 1440×960 , 2240×1488 , and 2304×1536 , respectively. We randomly selected 120 images (60 normal, 30 images with retinopathy grade 2, and 30 images with retinopathy grade 3) for each image resolution, to obtain a similar image distribution. The

TABLE 4. Confusion Matrices, Weighted κ Agreement, and 95% CI for Risk of Macular Edema between the Reference Standard and the Experts

	Expert A			Expert B		
	Grade 0	Grade 1	Grade 2	Grade 0	Grade 1	Grade 2
Grade 0	810	117	47	899	34	41
Grade 1	4	42	29	29	31	15
Grade 2	5	5	141	29	7	115
	$\kappa = 0.667$			$\kappa = 0.657$		
	95% CI = 0.623–0.710			95% CI = 0.605–0.709		

The columns indicate the experts' performance; the rows indicate the reference standard.

TABLE 5. Confusion Matrices, Weighted κ Agreement, and 95% CI for Retinopathy Grade between the Experts

	Grade 0	Grade 1	Grade 2	Grade 3
Grade 0	599	151	76	1
Grade 1	0	7	12	3
Grade 2	2	24	110	26
Grade 3	3	0	49	137
	$\kappa = 0.694$			
	95% CI = 0.664–0.725			

The columns indicate the performance of Expert B; the rows indicate the performance of Expert A.

TABLE 6. Confusion Matrices, Weighted κ Agreement and 95% CI for Risk of Macular Edema between the Experts

	Grade 0	Grade 1	Grade 2
Grade 0	778	21	20
Grade 1	114	30	20
Grade 2	65	21	131
$\kappa = 0.565$			
95% CI = 0.516–0.614			

The columns indicate the performance of Expert B; the rows indicate the performance of Expert A.

performance was not significantly different for images with different resolutions. Therefore, we cannot state that the CAD system performed better in images with higher resolution. However, more experiments should be performed to obtain a reliable measurement.

This study has limitations that need to be considered. First, the performance of the CAD is only compared to a single reading by two experts. Annotations from more specialists and the establishment of a gold standard are needed to perform a meaningful evaluation of the CAD's performance. Second, the CAD system does not detect isolated large hemorrhages. This may be the explanation for the slightly lower performance compared to expert A. The system could therefore be improved by adding a dedicated component for the identification of large hemorrhages. It should be noted that the proposed CAD system focused only in the detection of the earliest signs of DR, without a differentiation between different retinopathy grades. In future work, our research will be oriented to identify lesions that appear in more advance stages of DR, such as neovascularization, and to provide an automatic analysis of the retinopathy grade and the risk of macular edema.

Together with previous studies,^{5,15} this study confirms that the proposed CAD system reaches similar results in different databases and performance comparable to human observers. In addition, it has been shown that automated grading methods for DR screening is a cost-effective alternative to manual grading.⁸ In the view of these facts, the proposed CAD system is likely to be considered for screening practice, provided the remaining procedural, safety, and legal issues are resolved.

In conclusion, this study showed the performance of a comprehensive DR screening system on an independent, publicly available database. The performance of the system on this dataset is comparable to that of human experts and in accordance with the results obtained in previous studies. The system offers retinopathy screening programs a fast solution to reduce the burden of screening diabetes population while maintaining a high sensitivity.

References

- About diabetes: World Health Organization report, <http://www.who.int/diabetes/facts/en/index.html/>. Accessed March 31, 2009.
- Kinyoun J, Barton F, Fisher M, Hubbard L, Aiello L, Ferris F. Detection of diabetic macular edema: ophthalmoscopy versus photography—early treatment diabetic retinopathy study report number 5. The ETDRS research group, *Ophthalmology*. 1989;96:746–750.
- Early Treatment Diabetic Retinopathy Study Research Group, Early photocoagulation for diabetic retinopathy. ETDRS report 9. *Ophthalmology*. 1991;99:766–785.
- Bresnick GH, Mukamel DB, Dickinson JC, Cole DR. A screening approach to the surveillance of patients with diabetes for the presence of vision-threatening retinopathy, *Ophthalmology*. 2000;107(1):19–24.
- Abramoff MD, Niemeijer M, Suttorp-Schulten MSA, Viergever MA, Russell SR, van Ginneken B. Evaluation of a system for automatic detection of diabetic retinopathy from color fundus photographs in a large population of patients with diabetes. *Diabetes Care*. 2008;31(2):193–198.
- Lin DY, Blumenkranz MS, Brothers RS, Grosvenor DM. The sensitivity and specificity of single-field nonmydriatic monochromatic digital fundus photography with remote image interpretation for diabetic retinopathy screening: A comparison with ophthalmoscopy and standardized mydriatic color photography. *Am J Ophthalmol*. 2002;134:204–213.
- Williams GA, Scott IU, Haller JA, Maguire AM, Marcus D, McDonald HR. Single-field fundus photography for diabetic retinopathy screening: a report by the American Academy of Ophthalmology. *Ophthalmology*. 2004;111:1055–1062.
- Scotland GS, McNamee P, Philip S, et al. Cost-effectiveness of implementing automated grading within the national screening programme for diabetic retinopathy in Scotland. *Br J Ophthalmol*. 2007;91:1518–1523.
- Abramoff MD, Reinhardt JM, Russell SR. Automated early detection of diabetic retinopathy. *Ophthalmology*. 2010;117:1147–1154.
- Agurto C, Murray V, Barriga E, et al. Multiscale am-fm methods for diabetic retinopathy lesion detection. *IEEE Trans Med Imaging*. 2010;29(2):502–512.
- Teng T, Lefley M, Claremont D. Progress towards automated diabetic ocular screening: a review of image analysis and intelligent systems for diabetic retinopathy. *Med Biol Eng Comput*. 2002;40:2–13.
- Diaretdb1 v2.1. Diabetic retinopathy database and evaluation protocol, <http://www2.it.lut.fi/project/imageret/>. Accessed May 2, 2010.
- Retinopathy online challenge (ROC). <http://roc.healthcare.uiowa.edu/>. Accessed June 6, 2010.
- Methods to evaluate segmentation and indexing techniques in the field of retinal ophthalmology. <http://messidor.crihan.fr/>. Accessed September 6, 2010.
- Niemeijer M, Abramoff MD, van Ginneken B. Information fusion for diabetic retinopathy CAD in digital color fundus photographs, *IEEE Trans Med Imaging*. 2009;28(5):775–785.
- Niemeijer M, Abramoff MD, van Ginneken B. Image structure clustering for image quality verification of color retina images in diabetic retinopathy screening. *Med Image Anal*. 2006;10(6):888–898.
- Niemeijer M, Staal JJ, van Ginneken B, Loog M, Abramoff MD. Comparative study of retinal vessel segmentation methods on a new publicly available database, in medical imaging. *Proc SPIE*. 2004;5370:648–656.
- Niemeijer M, Abramoff MD, van Ginneken B. Fast detection of the optic disc and fovea in color fundus photographs. *Med Image Anal*. 2009;13(6):859–870.
- Niemeijer M, van Ginneken B, Staal J, Suttorp-Schulten, Abramoff MD. Automatic detection of red lesions in digital color fundus photographs. *IEEE Trans Med Imaging*. 2005;24(5):584–592.
- Niemeijer M, van Ginneken B, Russel SR, Suttorp-Schulten MS, Abramoff MD. Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis, *Invest Ophthalmol Vis Sci*. 2007;48:2260–2267.
- Metz CE, Herman BA, Roe CA. Statistical comparison of two roc-curve estimates obtained from partially-paired datasets. *Med Decis Making*. 1998;18(1):110–121.
- Metz CE. ROC methodology in radiologic imaging, *Invest Radiol*. 1986;21:720–733.